DOCUMENTO EXPLICATIVO DE LAS VARIABLES CONTENIDAS EN LA BASE DE DATOS

La generación de tres bases de datos, una para cada cantón de estudio, tiene como finalidad colectar información de variables relacionadas con el inicio de incendios forestales (momento de ignición), cuya inclusión en este proyecto surge de una revisión de literatura científica previa.

Estas bases de datos son el primer paso para generar los modelos de susceptibilidad de incendios forestales, que nos permiten identificar zonas con una probabilidad de ocurrencia mayor y que, por tanto, necesitan una atención prioritaria en la prevención de estos desastres.

Este documento surge en colaboración con tres estudiantes del componente práctico ***Gestión Productiva 3.1.*** de la titulación de ***Gestión Ambiental,*** y pretende dar una breve explicación del proceso de obtención de las ***23 variables*** que conforman las bases de datos anteriormente mencionadas. En función de ello, esta guía se subdivide en tres secciones: **variables topológicas, variables de interacción antrópica y variables del estado de la vegetación.**

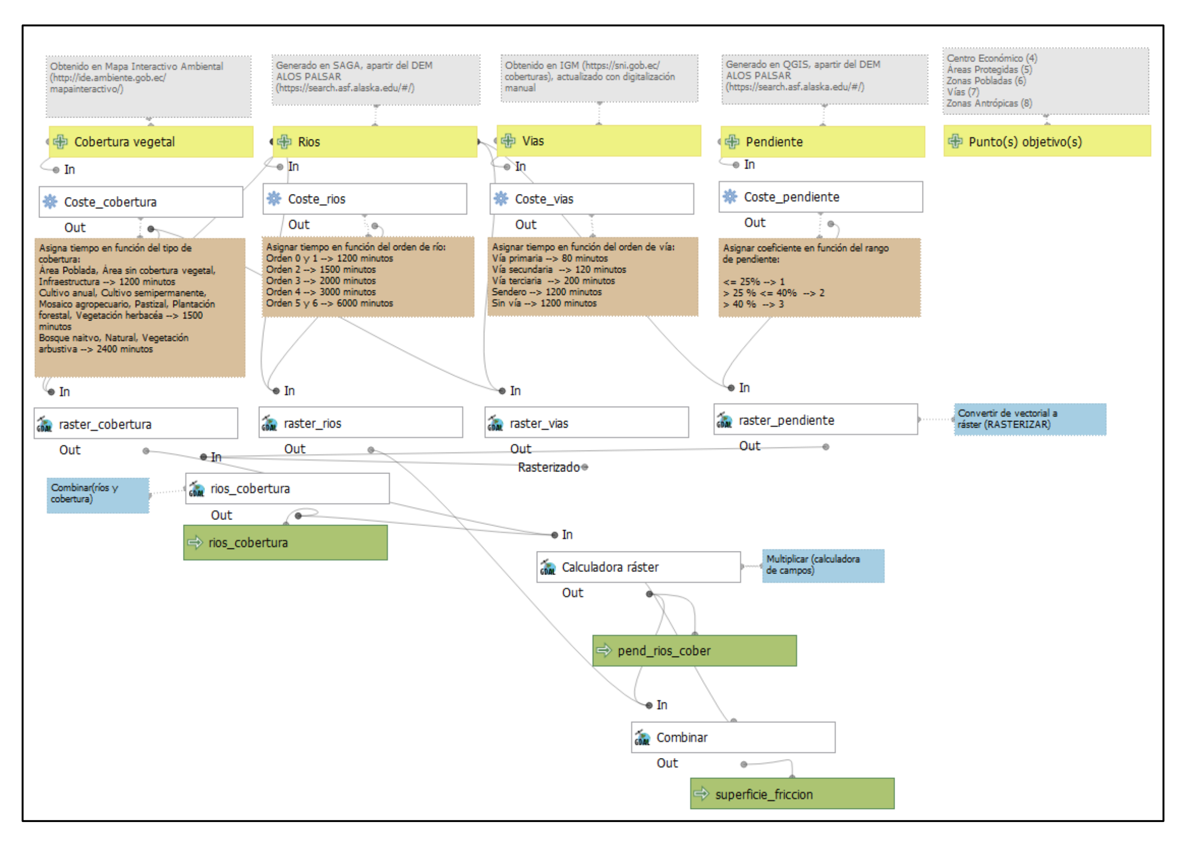
Puede acceder a las bases de datos en el siguiente enlace.

VARIABLES TOPOGRÁFICAS (https://bit.ly/3HTPxjx)

Variable		Descripción	Proceso de obtención	Metadatos de la capa
1	Elevación	Altura sobre el nivel del mar.	Se importa un archivo .kml con los límites del cantón de interés y se realiza la búsqueda de Modelos de Elevación Digital (DEM) disponibles en la web de ALOS PALSAR para la zona. Una vez descargados los archivos en *high resolution* se utiliza la **herramienta de QGIS** para combinar las teselas (Ráster → Miscelánea → Combinar), posteriormente se recorta a los límites de nuestra zona de interés (Ráster → Extracción → Cortar ráster por capa de máscara)	**SRC:** EPSG:32717 **Resolución espacial:** 12.5 x 12.5 metros **Generadores:** Joel Guillén Julia Loján Néstor Merino Esteban Proaño **Rango de valores:** 0 – 3668 **Unidades:** m.s.n.m.
2	Pendiente	Grado de inclinación del terreno, puede medirse en grados o porcentaje.	Al DEM obtenido en la web de ALOS PALSAR, se le aplica un filtro de suavizado en QGIS **herramienta SAGA** (Simple Filter, Smooth 3). Después de aplicar el filtro de suavizado, se utiliza la **herramienta GRASS** (r.slope.aspect, percent)	**SRC:** EPSG:32717 **Resolución espacial:** 12.5 x 12.5 metros **Generadores:** Joel Guillén Julia Loján Néstor Merino Esteban Proaño **Rango de valores:** 0 – 375 **Unidades:** porcentaje (%)
3	Orientación	Dirección de pendiente descendente que presenta cada celda, tomando como referencia los puntos cardinales (Norte, Sur, Este, Oeste)	Al DEM obtenido en la web de ALOS PALSAR, se le aplica un filtro de suavizado en QGIS **herramienta SAGA** (Simple Filter, Smooth 3). Después de aplicar el filtro de suavizado, se utiliza la **herramienta GRASS** (r.slope.aspect)	**SRC:** EPSG:32717 **Resolución espacial:** 12.5 x 12.5 metros **Generadores:** Joel Guillén Julia Loján Néstor Merino Esteban Proaño **Rango de valores:** 0 – 360 **Unidades:** grados (°)

VARIABLES DE INTERACCIÓN ANTRÓPICA (https://bit.ly/3HTPxjx)

La generación de las capas de accesibilidad surge a partir de una capa de ***superficie de fricción,*** que contiene el costo añadido que implica la movilización desde cada punto del territorio hasta el punto(s) objetivo(s), la siguiente gráfica pretende ilustrar el proceso seguido para su cálculo.



Metodología tomada de:
Reyes, F. (2017). Mapa de Accesibilidad Territorial. En Manual de Prácticas de Ordenación Territorial (págs. 29-41). Loja: UTPL.

Variable		Descripción	Proceso de obtención	Metadatos de la capa
4	Accesibilidad al centro económico	Facilidad de movimiento desde cualquier punto del cantón hasta el centro económico.	Utilizando la herramienta GRASS **r.cost.full,** se configura como punto(s) objetivo(s) la capa de centro económico descargada en el portal del SNI.	**SRC:** EPSG:32717 **Resolución espacial:** 10 x 10 metros **Generadores:** Joel Guillén Julia Loján Néstor Merino Esteban Proaño **Rango de valores:** 0 – 343 **Unidades:** minutos
5	Accesibilidad a áreas protegidas	Facilidad de movimiento desde cualquier punto del cantón hasta áreas protegidas.	Utilizando la herramienta GRASS **r.cost.full,** se configura como punto(s) objetivo(s) la capa de áreas protegidas descargada en el Mapa Interactivo Ambiental. Las categorías consideradas de áreas protegidas son: Sistema Nacional de Áreas Protegidas Bosque y Vegetación Protectora Área Bajo Conservación Patrimonio Forestal del Estado Áreas de Protección Hídrica	**SRC:** EPSG:32717 **Resolución espacial:** 10 x 10 metros **Generadores:** Joel Guillén Julia Loján Néstor Merino Esteban Proaño **Rango de valores:** 0 – 271 **Unidades:** minutos

6	Accesibilidad a zonas pobladas	Facilidad de movimiento desde cualquier punto del cantón hasta poblados.	Utilizando la herramienta GRASS **r.cost.full,** se configura como punto(s) objetivo(s) la capa de poblados descargada en el portal del SNI.	**SRC:** EPSG:32717 **Resolución espacial:** 10 x 10 metros **Generadores:** Joel Guillén Julia Loján Néstor Merino Esteban Proaño **Rango de valores:** 0 – 365 **Unidades:** minutos
7	Accesibilidad a vías	Facilidad de movimiento desde cualquier punto del cantón hasta vías.	Utilizando la herramienta GRASS **r.cost.full,** se configura como punto(s) objetivo(s) la capa de vías descargada en el portal del SNI y actualizada manualmente.	**SRC:** EPSG:32717 **Resolución espacial:** 10 x 10 metros **Generadores:** Joel Guillén Julia Loján Néstor Merino Esteban Proaño **Rango de valores:** 0 – 360 **Unidades:** minutos
8	Accesibilidad a zonas antrópicas	Facilidad de movimiento desde cualquier punto del cantón hasta zonas pobladas o zonas agropecuarias.	Utilizando la herramienta GRASS **r.cost.full,** se configura como punto(s) objetivo(s) la capa de cobertura vegetal descargada en el Mapa Interactivo Ambiental.	**SRC:** EPSG:32717 **Resolución espacial:** 10 x 10 metros **Generadores:** Joel Guillén Julia Loján Néstor Merino Esteban Proaño **Rango de valores:** 0 – 262 **Unidades:** minutos
9	Distancia desde el centro económico	Distancia desde el centro económico.	Utilizando la herramienta GRASS **r.grow.distance,** se configura como punto(s) objetivo(s) la capa de centro económico descargada en el portal del SNI.	**SRC:** EPSG:32717 **Resolución espacial:** 10 x 10 metros **Generadores:** Joel Guillén Julia Loján Néstor Merino Esteban Proaño **Rango de valores:** 0 – 4173 **Unidades:** metros
10	Distancia desde ríos	Distancia desde ríos.	Utilizando la herramienta GRASS **r.grow.distance,** se configura como punto(s) objetivo(s) la capa ríos generada a partir del DEM.	**SRC:** EPSG:32717 **Resolución espacial:** 10 x 10 metros **Generadores:** Joel Guillén Julia Loján Néstor Merino Esteban Proaño **Rango de valores:** 0 – 4173 **Unidades:** metros
11	Distancia desde vías	Distancia desde vías.	Utilizando la herramienta GRASS **r.grow.distance,** se configura como punto(s) objetivo(s) la capa vías descargada en el portal del SNI y actualizada manualmente.	**SRC:** EPSG:32717 **Resolución espacial:** 10 x 10 metros **Generadores:** Joel Guillén Julia Loján Néstor Merino Esteban Proaño **Rango de valores:** 0 – 14253 **Unidades:** metros
12	Distancia desde áreas protegidas	Distancia desde áreas protegidas	Utilizando la herramienta GRASS **r.grow.distance,** se configura como punto(s) objetivo(s) la capa áreas protegidas descargada en el Mapa Interactivo Ambiental.	**SRC:** EPSG:32717 **Resolución espacial:** 10 x 10 metros **Generadores:** Joel Guillén Julia Loján Néstor Merino Esteban Proaño

			Las categorías consideradas de áreas protegidas son: Sistema Nacional de Áreas Protegidas Bosque y Vegetación Protectora Área Bajo Conservación Patrimonio Forestal del Estado Áreas de Protección Hídrica	**Rango de valores:** 0 – 15047 **Unidades:** metros
13	Distancia a zonas antrópicas	Distancia desde zonas pobladas o zonas agropecuarias.	Utilizando la herramienta GRASS **r.grow.distance,** se configura como punto(s) objetivo(s) la capa zonas pobladas o zonas agropecuarias.	**SRC:** EPSG:32717 **Resolución espacial:** 10 x 10 metros **Generadores:** Joel Guillén Julia Loján Néstor Merino Esteban Proaño **Rango de valores:** 0 – 661 **Unidades:** metros

VARIABLES DEL ESTADO DE VEGETACIÓN (https://terabox.com/s/1WtFtvbjo_NW_VI1gYILWIA Password:t5kq)

Variable		Descripción	Proceso de obtención	Metadatos de la capa
14	ARVI	***Atmospherically Resistant Vegetation Index*** NDVI corregido para efectos de dispersión atmosférica $ARVI = \frac{NIR - (2 * Red) + Blue}{NIR + (2 * Red) + Blue}$ NIR (Near-Infrared Region) → B8 Red (Rojo, espectro visible) → B11 Blue (Azul, espectro visible) → B2	Los índices espectrales o índices de vegetación son descargados mediante el siguiente script de Google Earth Engine, se usa para ello imágenes Sentinel 2A.	**SRC:** EPSG:32717 **Resolución espacial:** 10 x 10 metros **Rango de valores:** -1 a 1
15	BSI	***Bare Soil Index*** Identificar zonas sin cobertura vegetal $BSI = \frac{(Red + SWIR) - (NIR + Blue)}{(Red + SWIR) + (NIR + Blue)}$ Red (Rojo, espectro visible) → B11 SWIR (Shortwave Infrared) → B4 NIR (Near-Infrared Region) → B8 Blue (Azul, espectro visible) → B2		**SRC:** EPSG:32717 **Resolución espacial:** 10 x 10 metros **Rango de valores:** -1 a 1
16	GNDVI	***Green Normalized Difference Vegetation Index*** Versión del NDVI más sensible a la variación del contenido de clorofila y nitrógeno en el cultivo $GNDVI = \frac{NIR - SWIR}{NIR + SWIR}$ SWIR (Shortwave Infrared) → B3 NIR (Near-Infrared Region) → B8		**SRC:** EPSG:32717 **Resolución espacial:** 10 x 10 metros **Rango de valores:** -1 a 1
17	MSI	***Moisture Stress Index*** Determinar el estrés hídrico de la vegetación $MSI = \frac{MidIR}{NIR}$ MidIR (Shortwave Infrared) → B11 NIR (Near-Infrared Region) → B8		**SRC:** EPSG:32717 **Resolución espacial:** 10 x 10 metros **Rango de valores:** 0 a 3+

18	NBR	***Normalized Burn Ratio*** Estimar la cantidad y desarrollo de áreas incendiadas $NBR = \frac{NIR - SWIR}{NIR + SWIR}$ SWIR (Shortwave Infrared) → B12 NIR (Near-Infrared Region) → B8		**SRC:** EPSG:32717 **Resolución espacial:** 10 x 10 metros **Rango de valores:** -1 a 1
19	NDMI	***Normalized Difference Moisture Index*** Determinar el contenido de agua de la vegetación $NDMI = \frac{NIR - SWIR}{NIR + SWIR}$ SWIR (Shortwave Infrared) → B11 NIR (Near-Infrared Region) → B8		**SRC:** EPSG:32717 **Resolución espacial:** 10 x 10 metros **Rango de valores:** -1 a 1
20	NDWI	***Normalized Difference Water Index*** Identificar masas de agua y zonas de elevada humedad $NDWI = \frac{NIR - Green}{NIR + Greem}$ SWIR (Shortwave Infrared) → B3 NIR (Near-Infrared Region) → B8		**SRC:** EPSG:32717 **Resolución espacial:** 10 x 10 metros **Rango de valores:** -1 a 1
21	NDVI	***Normalized Difference Vegetation Index*** Facilidad con la que el fuego se propaga en la vegetación $NBR = \frac{NIR - Red}{NIR + Red}$ Red (Rojo, espectro visible) → B4 NIR (Near-Infrared Region) → B8		**SRC:** EPSG:32717 **Resolución espacial:** 10 x 10 metros **Rango de valores:** -1 a 1
22	SAVI	***Soil Adjusted Vegetation Index*** Analizar la vegetación en etapas de crecimiento inicial o dispersa $NBR = \frac{(NIR - R)}{(NIR + R + L)} * (1 + L)$ NIR (Near-Infrared Region) → B8 R (Rojo, espectro visible) → B4 Factor de correción de la luminosidad del suelo → L		**SRC:** EPSG:32717 **Resolución espacial:** 10 x 10 metros **Rango de valores:** -1 a 1+
23	Moisture Index	***Normalized Difference Moisture Index*** Estimar el contenido de humedad de la vegetación $Moisture = \frac{NIR - SWIR}{NIR + SWIR}$ SWIR (Shortwave Infrared) → B11 NIR (Near-Infrared Region) → B8A		**SRC:** EPSG:32717 **Resolución espacial:** 10 x 10 metros **Rango de valores:** -1 a 1+

Base de datos (puntos de ocurrencia y no ocurrencia) para los cantones de estudio

Límites

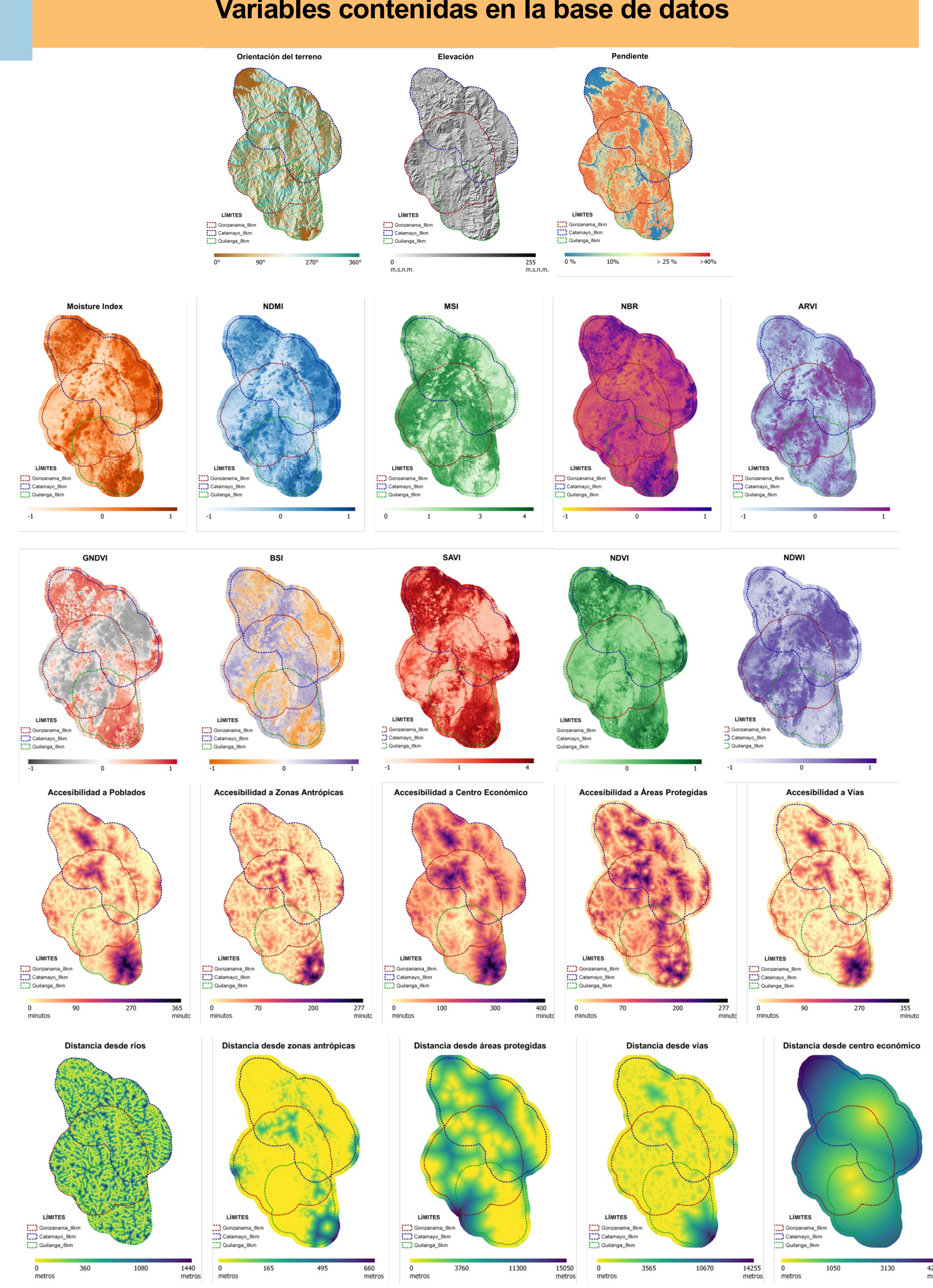
- Gonzanama (8km)
- Quilanga (8km)
- Catamayo (8km)

Puntos generados

- 1 - Ocurrencia
- 0 - Control

Cantón	Número puntos de NO ocurrencia	Número puntos ocurrencia
Quilanga	311	313
Catamayo	351	351
Gonzanama	427	432

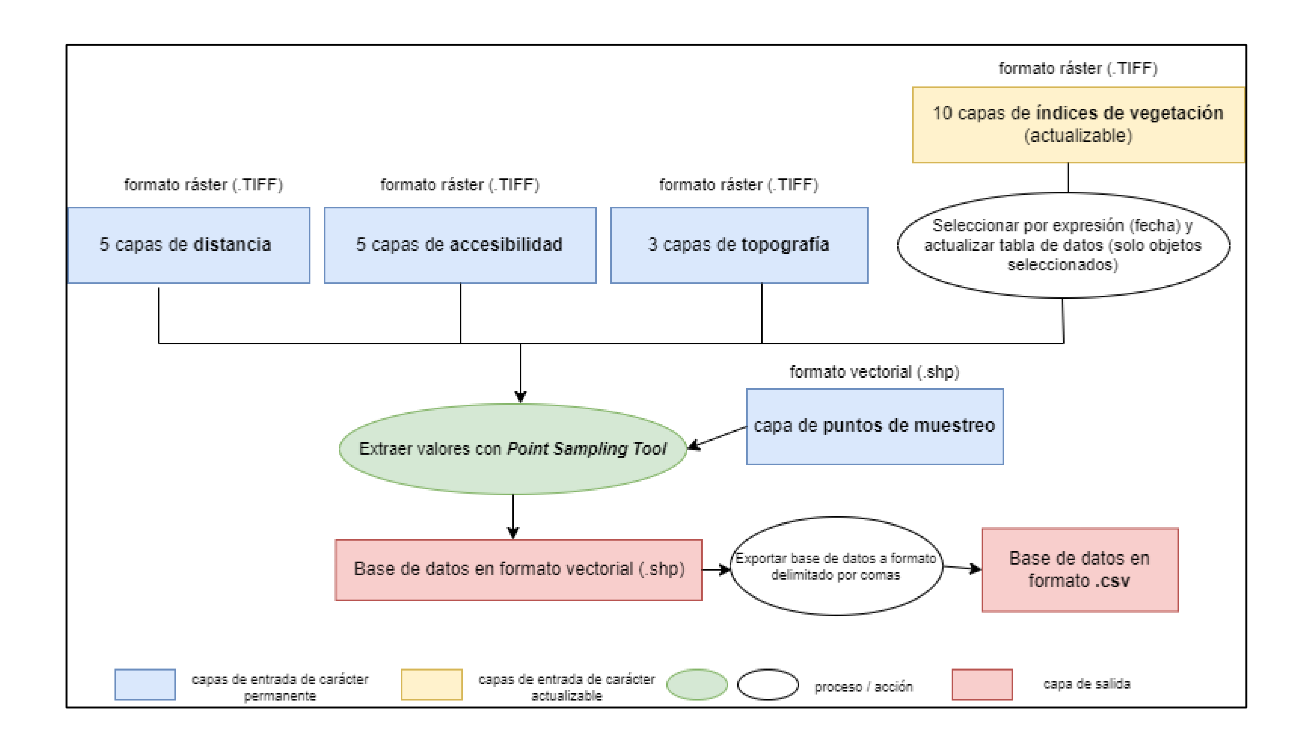
Variables contenidas en la base de datos



DOCUMENTO EXPLICATIVO DEL PROCESO DE GENERACIÓN DE LA BASE DE DATOS

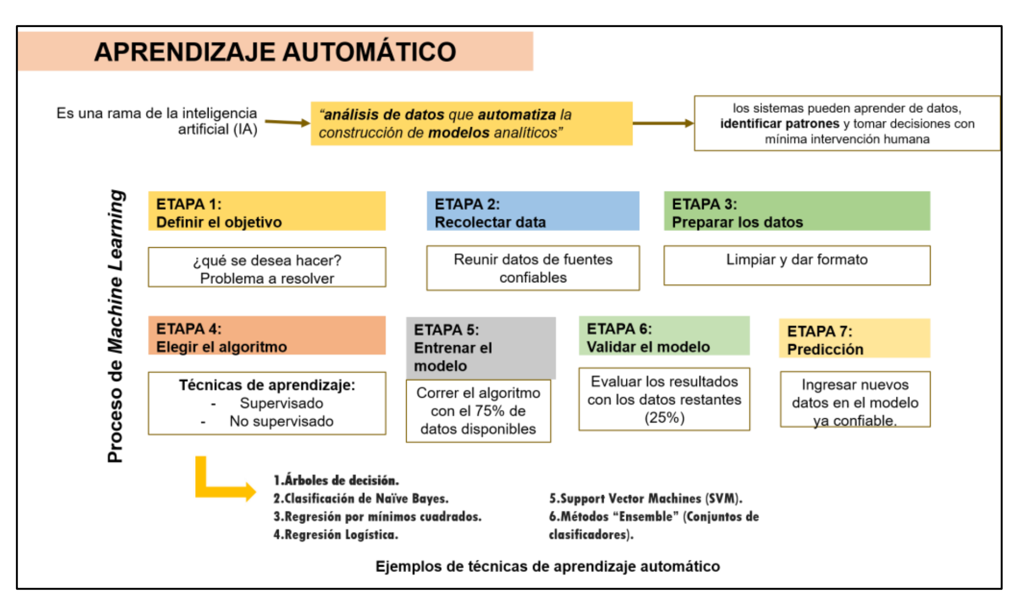
Una vez culminada la ***actividad 1*** dentro del ***objetivo específico 1*** que correspondía a la identificación y espacialización de las variables, se dio paso a la generación de las tres bases de datos. En este proceso se utiliza el software libre ***QGIS v.3.22.*** con su complemento ***Point Sampling Tool,*** el que permite capturar valores de capas ráster o vectoriales que le corresponden a los puntos de muestreo (puntos de ocurrencia y no ocurrencia mencionados en la evidencia 1).

De las 23 variables contenidas en las bases de datos, 13 tienen un carácter permanente y 10 son actualizables (disponibilidad de imágenes Sentinel 2A cada 5 días), por lo que el tratamiento de estos datos es distinto. Para ilustrar mejor este proceso se puede consultar el siguiente diagrama de flujo:



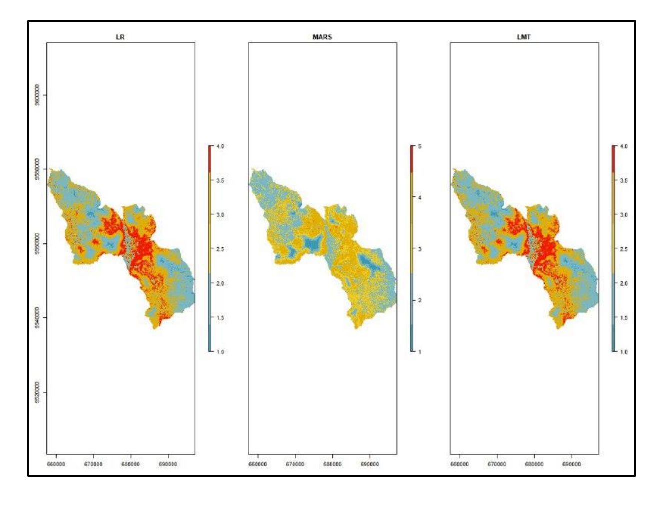
DOCUMENTO EXPLICATIVO DEL PROCESO DE GENERACIÓN DE LOS MODELOS DE SUSCEPTIBILIDAD

A partir de las bases de datos generadas en la ***actividad 2*** se inicia el proceso de generación y validación de modelos de aprendizaje automático o *machine learning*, en este caso aplicados a la identificación de zonas susceptibles a incendios forestales, la metodología seguida se describe en la siguiente gráfica:

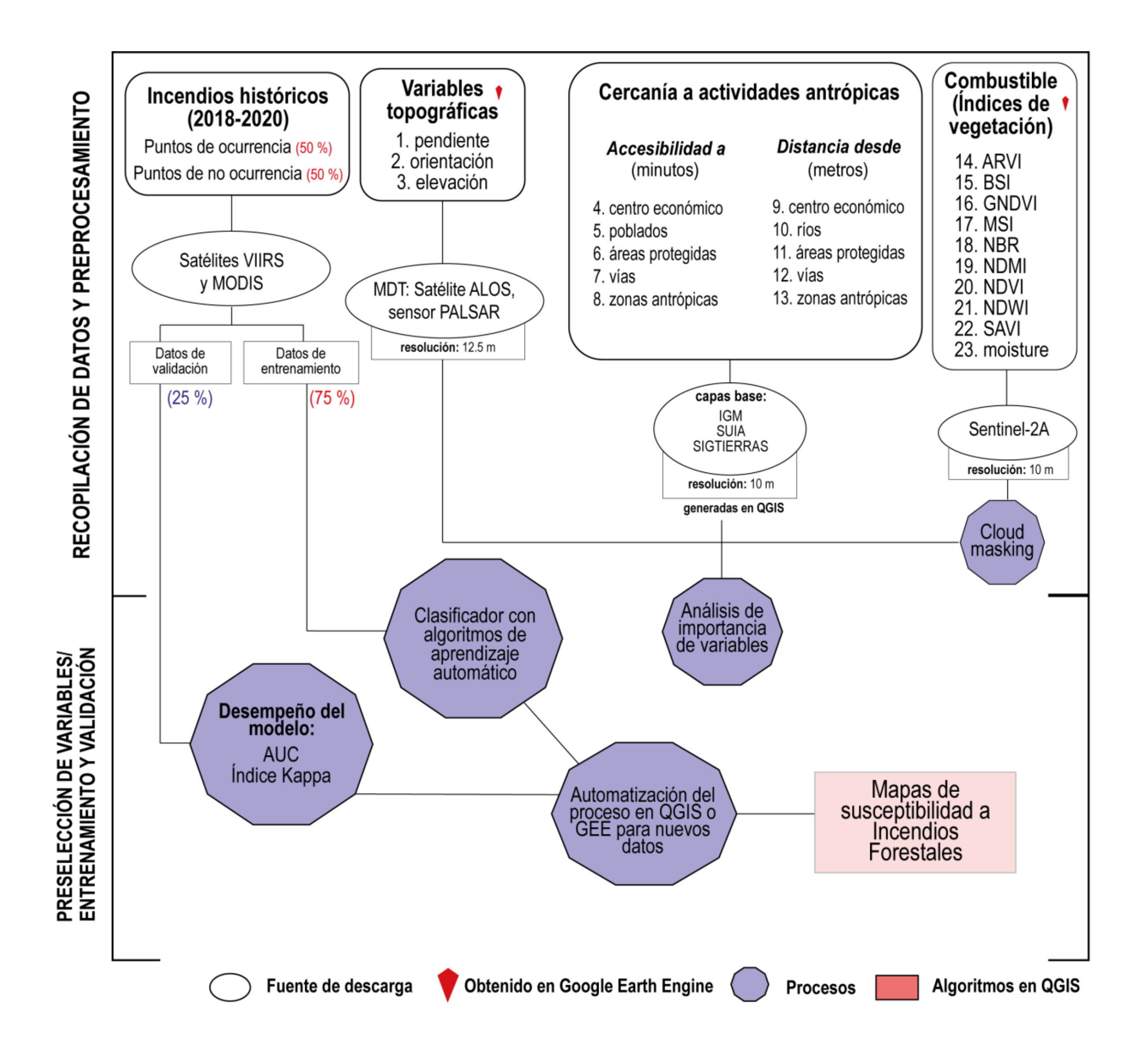


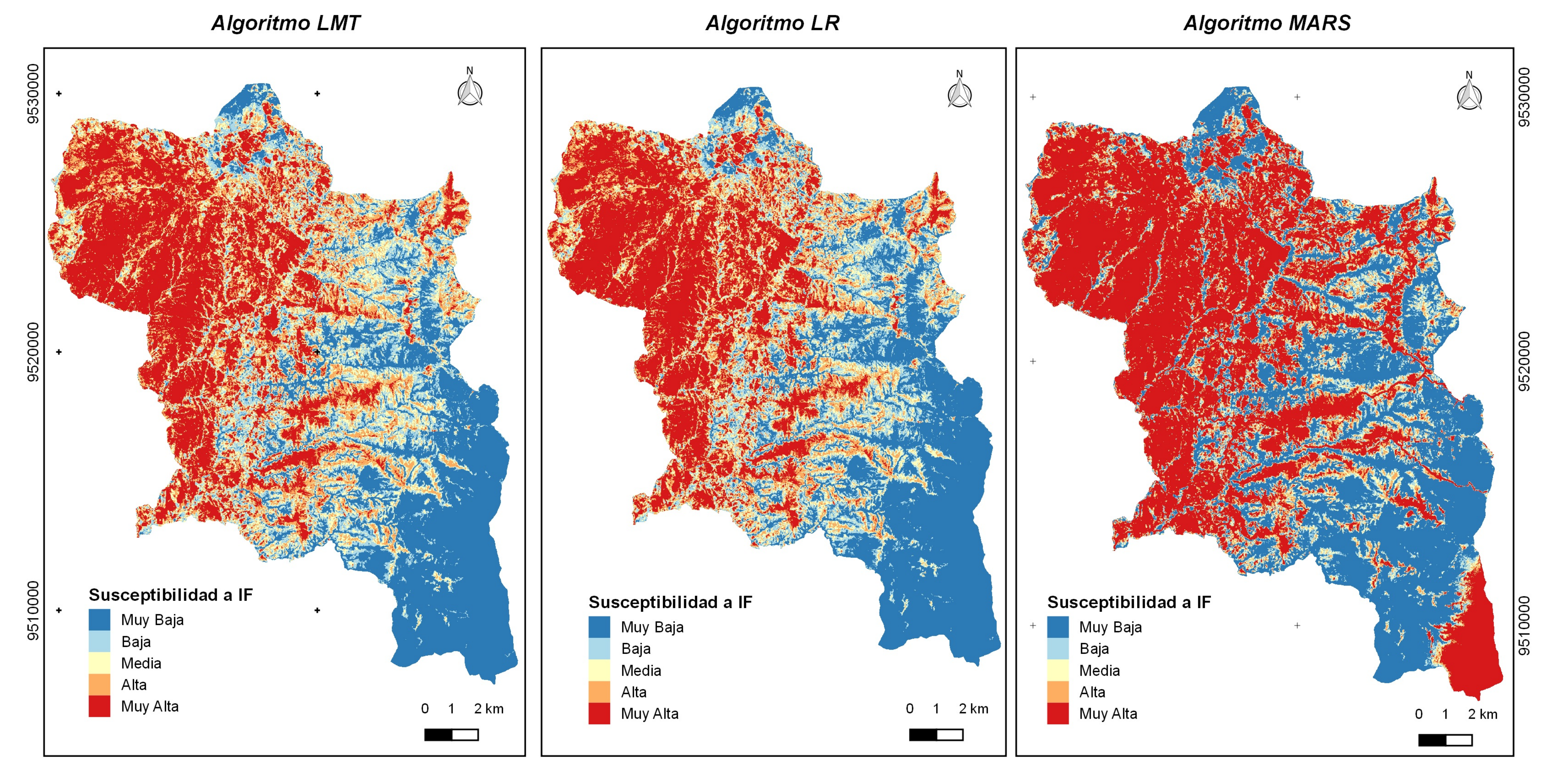
La ***actividad 3*** se ubica entre las ***etapas 3 y 4*** del proceso de *machine learning.* La ***etapa 1*** se abordó en la ***actividad 1*** mientras se identificaban las variables a utilizarse y las ***etapas 2 y 3*** se desarrollaron en la ***actividad 2*** con la herramienta ***Point Sampling Tool*** de QGIS. Las siguientes etapas (***5, 6, 7***) suceden en las ***actividades 4 y 5.***

La generación de los tres modelos de susceptibilidad (uno para cada cantón) siguió el proceso que se explica en la figura final de este documento. Se realiza una partición de las bases de datos en dos grupos: 75% para entrenar el modelo y 25% para validarlo según su desempeño en las métricas de evaluación. Para los tres cantones se probaron tres algoritmos: árboles de decisión logística (LMT), regresión logística (LR) y splines de regresión adaptativa multivariadas (MARS).

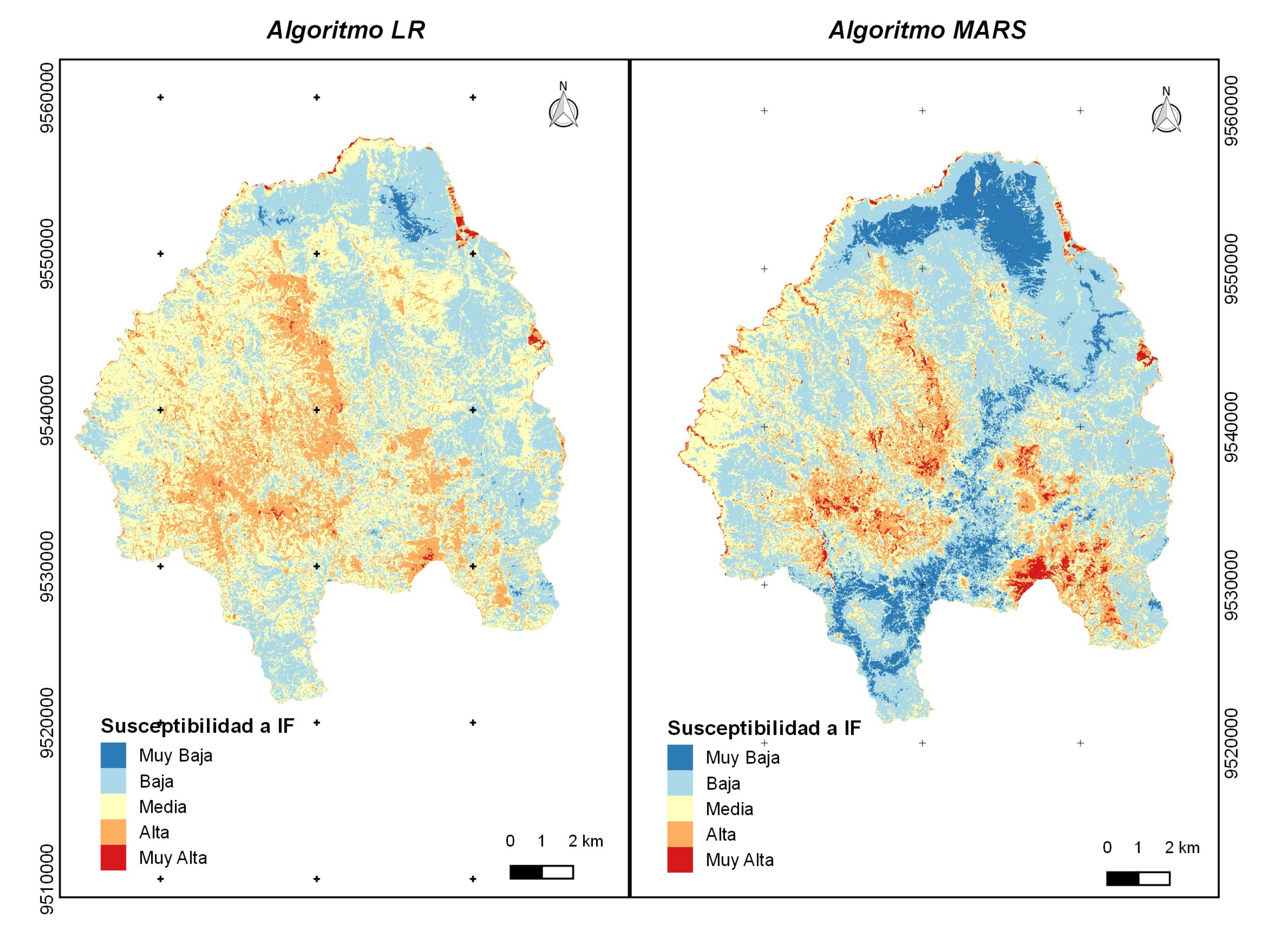


Ejemplo de la comparación de los tres algoritmos (cantón Catamayo)

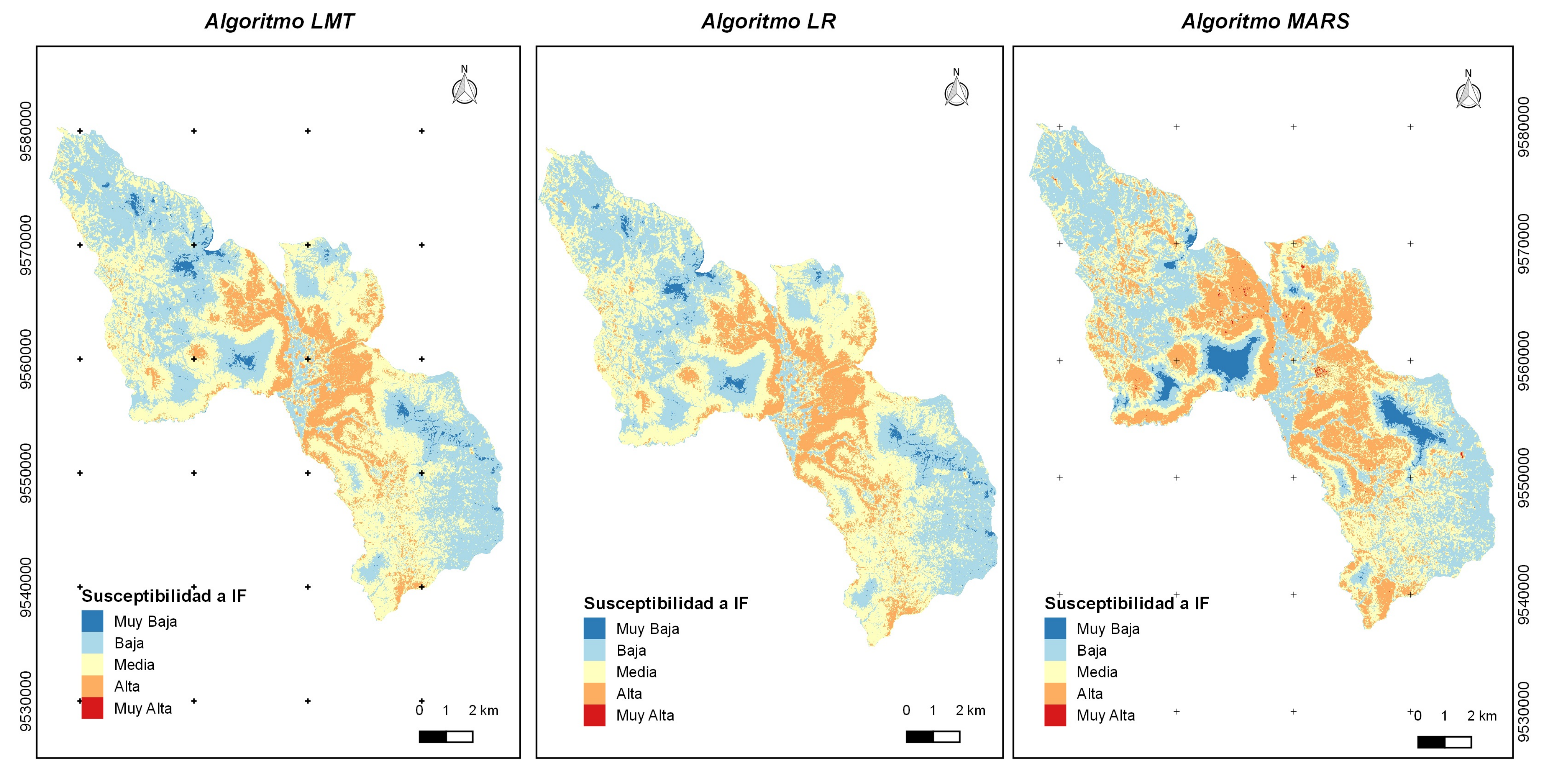

RECOPILACIÓN DE DATOS Y PREPROCESAMIENTO
Incendios históricos (2018-2020)
Puntos de ocurrencia (50 %)
Puntos de no ocurrencia (50 %)
Variables topográficas
1. pendiente
2. orientación
3. elevación
Cercanía a actividades antrópicas
Accesibilidad a (minutos)
4. centro económico
5. poblados
6. áreas protegidas
7. vías
8. zonas antrópicas
Distancia desde (metros)
9. centro económico
10. ríos
11. áreas protegidas
12. vías
13. zonas antrópicas
Combustible (Índices de vegetación)
14. ARVI
15. BSI
16. GNDVI
17. MSI
18. NBR
19. NDMI
20. NDVI
21. NDWI
22. SAVI
23. moisture
Satélites VIIRS y MODIS
MDT: Satélite ALOS, sensor PALSAR
resolución: 12.5 m
Datos de validación
Datos de entrenamiento
(25 %)
(75 %)
capas base: IGM SUIA SIGTIERRAS
resolución: 10 m
generadas en QGIS
Sentinel-2A
resolución: 10 m
Cloud masking
PRESELECCIÓN DE VARIABLES/ ENTRENAMIENTO Y VALIDACIÓN
Clasificador con algoritmos de aprendizaje automático
Análisis de importancia de variables
Desempeño del modelo:
AUC
Índice Kappa
Automatización del proceso en QGIS o GEE para nuevos datos
Mapas de susceptibilidad a Incendios Forestales
Fuente de descarga
Obtenido en Google Earth Engine
Procesos
Algoritmos en QGIS



MODELOS DE SUSCEPTIBILIDAD A INCENDIOS FORESTALES GENERADOS PARA EL CANTÓN QUILANGA



MODELOS DE SUSCEPTIBILIDAD A INCENDIOS FORESTALES GENERADOS PARA EL CANTÓN GONZANAMÁ



MODELOS DE SUSCEPTIBILIDAD A INCENDIOS FORESTALES GENERADOS PARA EL CANTÓN CATAMAYO

REPORTE DEL PROCESO DE VALIDACIÓN DE LOS MODELOS DE SUSCEPTIBILIDAD

Como se ha mencionado en los documentos de evidencias previas, a esta última actividad dentro del objetivo 1 del proyecto le corresponden las ***etapas 6 y 7*** del proceso de aprendizaje automático. En estas instancias, se evalúan los resultados con los datos de validación (25% del conjunto de datos total) y, si las métricas son buenas, se procede a la predicción con nuevos datos.

Para los tres cantones se probaron tres algoritmos de aprendizaje automático en el software R: árboles de decisión logística (LMT), regresión logística (LR) y splines de regresión adaptativa multivariadas (MARS). Independientemente, de la técnica o algoritmo empleado, estos fueron evaluados con dos herramientas: el *Área Bajo la Curva (AUC)* y el *Coeficiente Kappa de Cohen.*

1.1.1. Área Bajo la Curva

La curva ROC (*Receiver Operating Characteristive*) permite evaluar la capacidad del modelo entrenado para clasificar información. Este mide la relación entre la sensibilidad o "tasa de verdaderos positivos" y la 1-especificidad o "tasa de falsos positivos" (Fórmula 1 y 2).

$$Sensibilidad = \frac{VP}{(VP+FN)} \; ; VP = verdaderos\ positivos \;\; FN = falsos\ negativos \tag{1}$$

$$Especificidad = \frac{VN}{(FP+VN)} \; ; VN = verdaderos\ negativos \;\; FP = falsos\ positivos \tag{2}$$

Así, el área que se encuentre bajo esta curva representa el porcentaje de clasificaciones correctas. Generalmente, los valores superiores a 0.70 son considerados como buenos predictores (Cerda et al., 2012).

El AUC para los modelos entrenados en cada cantón luego de aplicar validación cruzada con grupos iterativos fue superior a 0.85; lo que significa que el modelo tiene un 85% de probabilidad para distinguir correctamente entre la posibilidad y la no-posibilidad de ocurrencia de un incendio.

1.1.2. Coeficiente Kappa de Cohen

El coeficiente Kappa o índice Kappa es un instrumento estadístico que estima la concordancia entre dos observadores o herramientas de medición para un mismo fenómeno, con el propósito de ajustar el efecto del azar en la proporción de la concordancia esperada, tal como establece la fórmula 3 (Landa et al., 2014).

$$Kappa = \frac{P_o - P_e}{1 - P_e} \; ; \tag{3}$$

$$P_o = proporción\ de\ concordancia\ observada$$

$$P_e = proporción\ de\ concordancia\ esperada\ por\ el\ azar$$

Para evaluar la concordancia de variables categóricas, especialmente aquellas con salida dicotómica o binaria (como en el presente estudio), Landis y Koch proponen la interpretación cualitativa más conocida. En ella, los valores menores a 0.40 representan una concordancia pobre o débil; entre 0.41 a 0.6, moderada; entre 0.61 a 0.8, concordancia buena o significante y aquellos valores mayores a 0.80 tienen un nivel de conformidad casi perfecto.

En los modelos probados para estos cantones el valor del índice Kappa rodea el 0.82. Siguiendo la línea explicativa anterior, se puede deducir que la conformidad entre la variable dependiente (ocurrencia) de la base de datos y las predicciones del modelo es casi perfecta.

Referencias Bibliográficas

Cerda, Jaime, & Cifuentes, Lorena. (2012). *Uso de curvas ROC en investigación clínica: Aspectos teórico-prácticos.* Revista chilena de infectología, 29(2), 138-141. https://dx.doi.org/10.4067/S0716-10182012000200003

Landa, F., Cruz, N., Coria, G., Rebolledo, G., & Hernández, S. (2014). *El Índice Kappa de Cohen como Herramienta para la Reducción de Ruido de Información Cerebral en la Identificación de Niveles de Valencia.* Puebla.